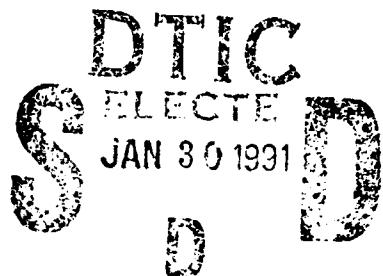


(2)

AD-A231 372

SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4022

DTIC COPY



Interior-Point Methods
for Convex Programming

by
Florian Jarre

TECHNICAL REPORT SOL 90-16[†]

November 1990

[†]This work was supported by a research grant from the Deutsche Forschungsgemeinschaft, and in part by the U.S. National Science Foundation Grant DDM-8715153 and the Office of Naval Research Grant N00014-90-J-1242.

Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do NOT necessarily reflect the views of the above sponsors.

Reproduction in whole or in part is permitted for any purposes of the United States Government. This document has been approved for public release and sale; its distribution is unlimited.

91 1 29 094

INTERIOR-POINT METHODS FOR CONVEX PROGRAMMING

Florian Jarre*

Technical Report SOL 90-16†
November 1990

Abstract

This work is concerned with generalized convex programming problems, where the objective and also the constraints belong to a certain class of convex functions. It examines the relationship of two conditions for generalized convex programming—self-concordance and a relative Lipschitz condition—and gives an outline for a short and simple analysis of an interior-point method for generalized convex programming. It generalizes ellipsoidal approximations for the feasible set, and in the special case of a nondegenerate linear program it establishes a uniform bound on the condition number of the matrices occurring when the iterates remain near the path of centers.

Key words: convex program, ellipsoidal approximation, relative Lipschitz condition, self-concordance

INTRODUCTION

In earlier papers, Jarre [4, 5], Mehrotra and Sun [9], and Nesterov and Nemirovsky [11, 12] tried to find a rather general class of convex programs that can be solved by interior-point methods. These authors use logarithmic barrier functions in their algorithms. Jarre and Mehrotra and Sun have imposed certain conditions on the constraint functions f_i , while Nesterov and Nemirovsky require the barrier function to be self-concordant. In all cases the conditions guarantee that Newton's method for minimizing the barrier function converges with a fixed rate of convergence.

When summarizing and relating some of the above results here, we attach great importance to the underlying geometry and structure of the method. To date, a large variety of interior-point methods and search directions have been suggested, all of which follow the same two components: centering and/or progress in the objective function. For the sake of clarity, only the method of centers is examined in detail to illustrate the geometry that is shared by all these methods and to form a foundation on which any of these methods can easily be analyzed. A short outline of how to derive a practical algorithm from the results presented here is given in Section 2.7.

*This work was supported by a research grant from the Deutsche Forschungsgemeinschaft, and in part by the U.S. National Science Foundation Grant DDM-8715153 and the Office of Naval Research Grant N00014-90-J-1242.

†On leave from Institut für Angewandte Mathematik, University of Würzburg, 8700 Würzburg, (West) Germany.



1. PROBLEM AND CONDITIONS

The problem under study is to find

$$\lambda^* := \min\{f_0(x) \mid x \in P\},$$

where

$$P := \{x \in \mathbb{R}^n \mid f_i(x) \leq 0 \quad \text{for } 1 \leq i \leq m\}, \quad (1.0)$$

and the $f_i \in C^2(P)$ are convex functions that fulfill certain conditions specified in Subsection 1.2. The first and second derivatives of $f_i(x)$ will sometimes be referred to as a row vector $Df_i(x)$ and a square matrix $D^2f_i(x)$, and sometimes as a linear form $Df_i(x)[\cdot]$ and a symmetric bilinear form $D^2f_i(x)[\cdot, \cdot]$.

For the sake of simplicity we assume that the interior of the feasible set P is nonempty and bounded. Given a point y in the intersection of the domains of the functions f_i , one can use a phase 1 algorithm as in the appendix of [3] to guarantee this assumption.

1.1. Possible Conditions on the f_i

1.1.1. Self-concordance

The most general condition is given in Nesterov and Nemirovsky [11], requiring that the barrier functions $\varphi_i(x) := -\ln(-f_i(x))$ are α -self-concordant on the interior P° of P for $1 \leq i \leq m$. Likewise, for $\lambda > \lambda^*$, the function $\varphi_0(x, \lambda) := -\ln(\lambda - f_0(x))$ is required to be self-concordant on P° .

Definition (self-concordance)

Here, in slight variation to the definition of [12], a function $\varphi : P^\circ \rightarrow \mathbb{R}$ is called self-concordant on P° with parameter α (in signs: $\varphi \in S_\alpha(P^\circ)$) if φ is three times continuously differentiable in P° and if for all $x \in P^\circ$ and all $h \in \mathbb{R}^n$ the following inequality holds:

$$|D^3\varphi(x)[h, h, h]| \leq 2\sqrt{\alpha}(D^2\varphi(x)[h, h])^{3/2}. \quad (1.1)$$

Intuitively, large values of α imply that the third derivative may be large, i.e. that φ cannot be well approximated by a quadratic function. Clearly, linear or convex quadratic functions fulfill (1.1) with a parameter $\alpha = 0$ on \mathbb{R}^n . However, we note that condition (1.1) is not applied to the constraint functions f_i themselves, but to the associated barrier functions φ_i (which are not linear or quadratic, even if the f_i are so).

For the sum $\varphi(x) = \sum_{i=1}^m \varphi_i(x)$, the following property is also required in [12].

Definition (strong self-concordance)

A function $\varphi : P^\circ \rightarrow \mathbb{R}$ is called strongly α -self-concordant (in signs: $\varphi \in S_\alpha^+(P^\circ)$) if it is α -self-concordant and if the level sets $\{x \in P^\circ \mid \varphi(x) \leq t\}$ are closed in \mathbb{R}^n for all $t \in \mathbb{R}$.

Intuitively this means that $\varphi(x)$ goes to infinity as x approaches the boundary

of P^o , a condition that is naturally fulfilled if the function φ is a penalty function defined as above, but may not hold for a general self-concordant function.

Remark 1 (Proposition 1.1 in [12])

The concept of self-concordance is affinely invariant in the following sense. If A is an invertible affine mapping, $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$, and φ is an α -self-concordant function, i.e. $\varphi \in S_\alpha(P^o)$, then the function $\tilde{\varphi}$ defined by $\tilde{\varphi}(x) := \varphi(A^{-1}x)$ is again α -self-concordant, $\tilde{\varphi} \in S_\alpha(AP^o)$.

Further the following simple rule for addition and scaling of self-concordant functions holds. If the functions φ_i are self-concordant with parameters α_i on the domains P_i^o for $i = 1, 2$ and if r_i are positive real numbers, then the function $\varphi := r_1\varphi_1 + r_2\varphi_2$ is α -self-concordant with $\alpha = \max\{\alpha_1/r_1, \alpha_2/r_2\}$ on the domain $P^o := P_1^o \cap P_2^o$.

Proof: Straightforward. ■

Note that Remark 1 also holds for the property of strong self-concordance.

1.1.2. Relative Lipschitz condition

The same motivation as above, having a function which is close to a quadratic function, has led to the definition of the following Relative Lipschitz Condition in Jarre [5]. (See also [2].) The functions f_i ($0 \leq i \leq m$) are supposed to be continuous on P and twice continuously differentiable functions on P^o , with Hessian matrices $D^2 f_i$ fulfilling the Relative Lipschitz Condition,

$$\begin{aligned} \exists M \geq 0 : \forall z \in \mathbb{R}^n \quad \forall y \in P^o \quad \forall h \text{ with } \|h\|_{H_i(y)} \leq 0.5/(1+M^{1/3}) \\ |z^T(D^2 f_i(y+h) - D^2 f_i(y))z| \leq M \|h\|_{H_i(y)} z^T D^2 f_i(y) z, \end{aligned} \quad (1.2)$$

which bounds the relative change of $D^2 f_i$ in neighboring points y and $y+h$ for small $\|h\|_{H_i(y)}$. Here, $\|\cdot\|_{H_i(y)}$ is a certain semi-norm that makes (1.2) affine invariant and is specified below. Again it is obvious that linear or convex quadratic functions f_i fulfill condition (1.2) with $M = 0$. (This condition is applied to the constraint functions f_i directly!) The precise definition of the matrix $H_i(y)$ and the associated semi-norm is given by

$$H_i(y) := D^2(-\ln(-f_i(x))) \Big|_{x=y} = \frac{D^2 f_i(y)}{-f_i(y)} + \frac{D^T f_i(y) D f_i(y)}{f_i^2(y)}$$

and $\|h\|_{H_i(y)}^2 := h^T H_i(y) h$. The matrices $H_i(y)$ arise as the Hessians of the logarithmic barrier functions. As shown below, the norm given by $H(y) := \sum_{i=1}^m H_i(y)$ is closely related to the shape of the feasible set P , and is a very convenient measure for analyzing Newton's method. Clearly $H_i(y)$ is positive semidefinite, since f_i is assumed to be convex. Note that condition (1.2) requires that $D^2 f_i(y+h)$ exists for all h with $\|h\|_{H_i(y)} \leq 0.5/(1+M^{1/3})$.

Remark 2

If condition (1.2) holds for the function $D^2 f_i$ at a point $y \in P^o$ (i.e. $f_i(y) < 0$) then also $f_i(y+h) < 0$ for all h with $\|h\|_{H_i(y)} \leq 0.5/(1+M^{1/3})$.

Proof: See Appendix. ■

Given a strictly feasible point y , i.e. a point y such that $f_i(y) < 0$ for all $1 \leq i \leq m$, condition (1.2) is only needed for points $y + h$ with $\|h\|_{H_i(y)} \leq 0.5/(1 + M^{1/3})$ for all $1 \leq i \leq m$. Remark 2 guarantees that also $f_i(y + h) < 0$ for all i . Hence $y + h \in P^\circ$, so that in fact condition (1.2) is needed only for points $y, y + h \in P^\circ$.

Example

The Relative Lipschitz Condition allows certain singularities on the boundary of P , the second derivative of the function $f : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto -\sqrt{x}$ e.g. fulfills the condition with $M = 8$ on $P := \{x|x \geq 0\}$.

1.1.3. Relationship between the relative Lipschitz condition and self-concordance

Loosely speaking, the Relative Lipschitz Condition is sufficient for the resulting barrier function to be self-concordant. More precisely one can state the following.

Lemma 1

If the second derivative $D^2 f$ of a convex function f fulfills the Relative Lipschitz Condition (1.2) (for infinitesimal $\|h\|$) on the domain $P_f := \{x|f(x) \leq 0\}$ and if f is three times continuously differentiable on P_f° , then the barrier function $\varphi(x) := -\ln(-f(x))$ is α -self-concordant on P_f° with the parameter $\alpha = (1 + M)^2$.

Proof: See Appendix. ■

The converse of Lemma 1 is not true; there even exist non-convex functions f whose barrier functions $\varphi(x) := -\ln(-f(x))$ are α -self-concordant (and hence convex) on P_f° (see Subsection 2.7. "Extensions"). The idea of self-concordance and Relative Lipschitz condition however are closely related, and as the following two statements show, self-concordance in fact is equivalent to a modified Relative Lipschitz condition. Lemma 2 is taken from [12].

Lemma 2 (Theorem 1.1 in [12])

Let φ be strongly α -self-concordant, $\varphi \in S_\alpha^+(P^\circ)$, and let a strictly feasible point $y \in P^\circ$ be given, and $h, z \in \mathbb{R}^n$. Define $H(y) := D^2\varphi(y)$, $\delta := \sqrt{h^T H(y) h} = \|h\|_{H(y)}$, and $x := y + h$.

Then the following is true: If $\delta < 1/\sqrt{\alpha}$, then

$$x = y + h \in P^\circ$$

and

$$(1 - \sqrt{\alpha}\delta)\|z\|_{H(y)} \leq \|z\|_{H(x)} \leq \frac{1}{(1 - \sqrt{\alpha}\delta)}\|z\|_{H(y)}.$$

Proof: See Appendix. ■

For $\delta \leq \frac{1}{2\sqrt{\alpha}}$ one has $\frac{1}{(1 - \sqrt{\alpha}\delta)^2} \leq 1 + 6\sqrt{\alpha}\delta$, and thus Lemma 2 implies that

$$|z^T(D^2\varphi(y + h) - D^2\varphi(y))z| \leq 6\sqrt{\alpha}\|h\|_{H(y)}z^T D^2\varphi(y)z$$

(cf. [3] (Lemma 2.1, equivalence of the H -norms)). Hence, a self-concordant barrier function φ also fulfills a Relative Lipschitz condition, where the norm of the vector

h is measured by $D^2\varphi(x)$ directly (and not by $D^2 - \ln(-\varphi(x))$). Conversely, it is easy to show the following

Remark 3

Let the function φ fulfill a Relative Lipschitz condition of the following form (with the notation of Lemma 2):

$$|z^T(D^2\varphi(x+h) - D^2\varphi(x))z| \leq \beta \|h\|_{H(x)} z^T D^2\varphi(x) z.$$

Then φ is self-concordant with parameter $\alpha = \beta^2/4$.

Proof: See Appendix. ■

1.1.4. Curvature constraint

Mehrotra and Sun [9] do not need the continuity of the second derivative of the functions f_i , but only a curvature constraint of the form

$$\exists \kappa \geq 1 : \forall h \in \mathbb{R}^n \forall x, y \in P : 0 < \kappa^{-2} h^T D^2 f_i(y) h \leq h^T D^2 f_i(x) h \leq \kappa^2 h^T D^2 f_i(y) h.$$

With this condition they can show the same result as Jarre and Nesterov and Nemirovsky for their algorithms. However, since in the above form the curvature constraint excludes linear or semidefinite quadratic functions f_i , as well as singularities on the boundary of P , we will not use this condition here.

1.2. Further Assumptions

In the following we will assume that the functions $-\ln(-f_i(x))$ are self-concordant with parameters α_i . Note that (by Lemma 1) the logarithmic barrier-functions φ_i of linear and convex quadratic functions f_i are 1-self-concordant, and so is their sum $\varphi = \sum_{i=1}^m \varphi_i(x)$ (by Remark 1). Thus, the following analysis includes linearly or quadratically constrained convex programming as a special case with $\alpha = 1$.

Without loss of generality we further assume that f_0 is linear. (Otherwise we may introduce an additional variable x_{n+1} , an additional constraint $f_{m+1}(x, x_{n+1}) := f_0(x) - x_{n+1} \leq 0$, and minimize x_{n+1} . Note, that for this construction the new function $-\ln(-f_{m+1})$ must be self-concordant on the domain $\{(x, x_{n+1}) | x \in P^o, f_0(x) < x_{n+1}\}$. In a practical implementation such a construction may increase the condition numbers of the Hessians considered in the algorithm.)

Note that by construction the resulting function $\varphi(x) = \sum_{i=1}^m \varphi_i(x)$ is *strongly* self-concordant on P^o .

2. PROPERTIES AND A SIMPLE METHOD

For $\lambda > \lambda^*$, let $P(\lambda)$ denote the feasible set P constrained by the additional inequality $f_0(x) = c^T x \leq \lambda$:

$$P(\lambda) := P \cap \{x \mid f_0(x) \leq \lambda\}.$$

The method outlined in this section follows a homotopy path $\lambda : \infty \rightarrow \lambda^*$ of some interior point $x(\lambda)$ in $P(\lambda)$. Here, $x(\lambda)$ is chosen as the well known *analytic center* of $P(\lambda)$ (Sonnevend [14]).

2.1. The Analytic Center

For each parameter $\infty \geq \lambda > \lambda^*$ the analytic center $x(\lambda)$ of $P(\lambda)$ is defined as the unique point x in $P(\lambda)^\circ$ minimizing the strictly convex logarithmic barrier function¹ φ :

$$\varphi(x, \lambda) := -q \ln(\lambda - f_0(x)) - \sum_{i=1}^m \ln(-f_i(x)) \quad (2.1)$$

with some fixed $q \in \mathbb{N}$ (the positive natural numbers) and $P(\lambda = \infty) := P$. In this paper only the choice $q = m$ is considered; the modification to other values of q is straightforward. The analytic center depends smoothly on all constraints, also on λ , and as the following analysis shows it can be efficiently approximated by Newton's method. The strict convexity of φ follows immediately from the boundedness of P and the *strong* self-concordance of φ on $P(\lambda)^\circ$.

The analytic center $x(\lambda)$ also maximizes the concave function of x

$$\Psi(x, \lambda) := \left[(\lambda - f_0(x))^q \prod_{i=1}^m (-f_i(x)) \right]^{1/(m+q)} \quad (2.2)$$

over P . One may interpret (2.2) as $x(\lambda)$ maximizing the product of the 'distances' to the constraints $f_i(x) \leq 0$.

Proof of concavity of (2.2): see appendix. ■

The analytic center $x(P)$ of a set P (or of the set $P(\lambda)$) is *invariant* under affine transformations of P in the sense that an invertible affine transformation $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ applied to the set P , $P \rightarrow AP = \{x \mid f(A^{-1}x) \leq 0\}$ also maps the analytic center $x(P) = \arg \max[\prod_{i=1}^m (-f_i(x))]^{1/m}$ to $Ax(P) = \arg \max[\prod_{i=1}^m (-f_i(A^{-1}x))]^{1/m} = x(AP)$. It is also *invariant* under scaling of the functions f_i .

The function φ in (2.1) is α -self-concordant on $P^\circ(\lambda)$ if the functions $\ln(-f_i(x))$ are α_i -self concordant, and, according to remark 1, $\alpha = \max\{\frac{1}{q}, \alpha_i, 1 \leq i \leq m\}$. Hence, for linear or quadratic f_i we have $\alpha = 1$ (by Lemma 1).

¹The function $\varphi(x, \lambda)$ in (2.1) defines the analytic center of $P(\lambda)$. For brevity we will also sometimes deal with the function $\varphi = -\sum_{i=1}^m \ln(-f_i(x))$ defining the analytic center of P . Similarly with $H(x) := D^2\varphi(x)$ and $H(x, \lambda) := D^2\varphi(x, \lambda)$. Results for $\varphi(x)$ and $H(x)$ are applied later to $\varphi(x, \lambda)$ and $H(x, \lambda)$.

2.2. Ellipsoidal Approximations of P

If the feasible set P is bounded, then the semi-norm in Lemma 2 is in fact a norm that is closely related to the geometrical shape of P . Lemma 2 already stated an inner ellipsoidal approximation of P ; for any point $x \in P^\circ$ the point $x + h \in P^\circ$ if h belongs to the ellipsoid defined by

$$\|h\|_{H(x)} \leq 1/\sqrt{\alpha}$$

where $H(x) = D^2\varphi(x)$. Furthermore one can show the following outer ellipsoidal approximation of P centered at its analytic center.

Lemma 3 (cf. [5] Corollary 2.15)

Let \bar{x} be the analytic center of P and $h \in \mathbb{R}^n$ be arbitrary with

$$\|h\|_{H(\bar{x})} \geq 16\sqrt{\alpha}m.$$

Then $\bar{x} + h \notin P$. **Proof:** See Appendix. ■

This two-sided ellipsoidal approximation of the feasible set P around its analytic center has been shown in [14] for the linear case and in [15, 3] for quadratic f_i (see also [5]). It relates the matrix H to the shape of the set P . In the next subsection we will show that the underlying norm $\|\cdot\|_H$ is also suitable when analyzing Newton's method.

2.3. Newton's Method

In the following we will give a proof of quadratic convergence of Newton's method for approximating the analytic center \bar{x} of a set P and give explicit constants (depending only on α) that describe the speed of convergence. Here all "distances" are measured in the H -norm and related to the concordance parameter α . Lemma 4 has been proved in modified form in [12] and states that if a Newton step for finding the center is small, then Newton's method converges. Conversely, Remark 4 guarantees that if a point y is sufficiently "close" to the analytic center \bar{x} of P , then again Newton's method converges. Recalling some notation, the Newton step $h(y)$ starting at y for finding the analytic center \bar{x} is given by $h(y) = -H^{-1}(y)D\varphi(y)^T$, with $H(y) = D^2\varphi(y)$.

Lemma 4 ([12], Theorem 1.3; quadratic convergence with constant $\frac{16}{9}\sqrt{\alpha}$)

Let φ be a strongly α -self-concordant function defined on a nonempty bounded set P° . For a point $y \in P^\circ$ define $H(y) := D^2\varphi(y)$, $g(y) := D\varphi(y)^T$. Let $h := h(y) = -H(y)^{-1}g(y)$ be the Newton step starting at y for finding the analytic center \bar{x} of P , let \tilde{h} be the following Newton step starting at $y + h$, and define the lengths of the Newton steps by $\delta := \|h\|_{H(y)}$ and $\tilde{\delta} := \|\tilde{h}\|_{H(y+h)}$. If

$$\delta < \frac{1}{\sqrt{\alpha}},$$

then $y + h$ is feasible, $y + h \in P^o$ and the length $\tilde{\delta}$ of the following Newton step is of order δ^2 :

$$\tilde{\delta} \leq \frac{\sqrt{\alpha}}{(1 - \sqrt{\alpha}\delta)^2} \delta^2.$$

For $\delta < \frac{1}{2\sqrt{\alpha}}$ this implies convergence of Newton's method, and for $\delta \leq \frac{1}{4\sqrt{\alpha}}$ it implies that $\tilde{\delta} \leq \frac{16}{9}\sqrt{\alpha}\delta^2$.

Proof: See Appendix. ■

The importance of this lemma is that the constant $\frac{16}{9}\sqrt{\alpha}$ can be explicitly stated, depending only on α and not on the data of the functions f_i .

For suitably damped Newton steps it is also shown in [12] that a fixed rate of convergence holds for the case $\delta > 1/4\sqrt{\alpha}$.

In the following subsections the result above will be applied to the function $\varphi(x, \lambda)$ (and $h(x, \lambda), H(x, \lambda)$) to analyze a "short-step" method for following the path of centers. A "long-step" method for convex constraint functions whose Hessians fulfill the Relative Lipschitz condition is analyzed in [2].

Remark 4

In the notation of the previous lemma, the following statement holds: If the length $\delta = \|h\|_{H(y)}$ of the Newton step h starting at y for finding the analytic center \bar{x} of P fulfills $\delta \leq \frac{1}{9\sqrt{\alpha}}$, then the "distance" from $y + h$ to \bar{x} is of the order δ^2 ; more precisely, $\|y + h - \bar{x}\|_{H(y)} \leq \frac{3}{2}\sqrt{\alpha}\delta^2$.

Proof: See Appendix. ■

These properties show that the length of the Newton step (in the H -norm) is a measure for the closeness to the center that can be used to analyze a method. In this context let us state two further remarks that are not needed for the analysis here but may be interesting for step-length control in a numerical implementation.

Remark 5

The H -norm of the Newton step $h(y)$ starting at a point $y \in P^o$ for finding the center \bar{x} of P is uniformly bounded for any $y \in P^o$ by $\|h(y)\|_{H(y)} \leq \sqrt{m}$. This bound does not depend on α . A similar observation is made in Proposition 3.5 in [12]. There are examples where $\|h(y)\|_{H(y)} \in O(\sqrt{m})$; see e.g. [13].

Proof: See Appendix. ■

Remark 6

Let some ϵ with $0 < \epsilon \leq \frac{1}{4}$ be given. If a point y satisfies $\|y - \bar{x}\|_{H(\bar{x})} \leq \frac{\epsilon}{\sqrt{\alpha}}$ then y lies in the domain of quadratic convergence of Newton's method and the Newton successor $y' = y - H(y)^{-1}h(y)$ satisfies

$$\|y' - \bar{x}\|_{H(\bar{x})} \leq \frac{1}{\sqrt{\alpha}} \frac{\epsilon^2}{(1 - \epsilon)^3}.$$

This is particularly interesting, since by Lemma 3 the whole set P is contained in a fixed "multiple" of this domain, namely $P \subset \{y \mid \|y - \bar{x}\|_{H(\bar{x})} \leq 16m\sqrt{\alpha}\}$.

Proof: See Appendix. ■

2.4. Short-Step Algorithm

Below, a short step algorithm is stated. This algorithm is too slow for a practical implementation. Its rate of convergence however ensures polynomiality in the case of a linear program (since the exact solution can be rounded from a sufficiently accurate approximation [1]). Further, the same rate of convergence can be guaranteed for a convex program with constraints whose logarithmic barrier functions are strongly self-concordant. Possible acceleration techniques for the algorithm that are based on the theoretical results developed here are outlined in Subsection 2.7. Implementations are discussed e.g. in [6, 8, 10, 2].

Under the assumptions of Sections 1 and 1.2 let a point $y_0 \in P^o$ and some number $\lambda_0 > \lambda^*$ be given such that the first Newton step $h(y_0, \lambda_0) \leq 1/(4\sqrt{\alpha})$. Simple modifications of the algorithm to generate such a point y_0 and λ_0 are omitted here (see e.g. [3]). Again, the objective function $f_0(x)$ is denoted by $c^T x$.

Algorithm

1. $k := 0; \sigma := 1/(8\sqrt{m}\sqrt{\alpha}); \epsilon = \text{desired accuracy}.$
2. $y_{k+1} := y_k + h_k$ where $h_k := h(k, \lambda_k) = -D^2\varphi(y_k, \lambda_k)^{-1}D\varphi(y_k, \lambda_k)^T$.
3. If $\lambda_k - c^T y_{k+1} \leq \frac{1}{2}\epsilon$ stop.
4. $\lambda_{k+1} := \lambda_k - \sigma(\lambda_k - c^T y_{k+1}).$
5. $k := k + 1;$ go to 1.

2.5. Convergence Analysis

In order to ensure convergence² of the algorithm the following two properties are shown.

First, after the update of λ_{k+1} in step 4, the iterate y_{k+1} again satisfies

$$\|h(y_{k+1}, \lambda_{k+1})\|_{H(y_{k+1}, \lambda_{k+1})} \leq 1/(4\sqrt{\alpha}). \quad (2.3)$$

This guarantees that the iterates remain feasible and close to the center.

Second,

$$\lambda_k - c^T y_{k+1} \geq \frac{17}{23}(c^T y_k - \lambda^*), \quad (2.4)$$

so, that the stopping criterion in step 3 is exact, and the gap $\lambda_k - \lambda^*$ in between the upper bound λ_k for $c^T y_k$ and the (unknown) optimal value λ^* is reduced by a factor of at least 0.4σ in step 4.

Proof: See Appendix. ■

This completes the proof of convergence!

²Only feasibility and convergence of the objective function value $c^T y_k$ to the optimal value are ensured.

2.6. Bounded Condition Numbers

Implementations of the affine scaling algorithm for solving linear programs encounter nearly singular Hessians if large step-lengths are chosen such that the iterates lie very near to the boundary of P . The following lemma shows that for nondegenerate linear programs this difficulty can be partly eliminated if the iterates remain in a neighborhood of the path of centers. Implementations in [6, 10, 8] show that with extrapolation techniques it is possible to generate fast algorithms that remain in such a neighborhood.

Lemma 5 (Estimate of worst-case condition numbers for the matrices $H(y)$ for nondegenerate linear problems)

Consider a (primal) nondegenerate linear program and any algorithm generating a sequence of points y_k in a σ -neighborhood of the path of centers with $\sigma = \frac{1}{4\sqrt{\alpha}}$. Here a point y is in the σ -neighborhood of the path of centers if the Newton step h starting at y for finding the “nearest” center measured in the H -norm is less than σ , i.e.,

$$\bar{\lambda} := \arg \min_{\lambda \geq 0} \|D\varphi(y, \lambda)\|_{H(y, \lambda)^{-1}}$$

defines the “nearest” center $x(\bar{\lambda})$ to y , and the corresponding Newton step

$$h := h(y, \bar{\lambda}) = -D^2\varphi(y, \bar{\lambda})D\varphi(y, \bar{\lambda})^T$$

satisfies $\|h\|_{H(y, \lambda)} \leq \sigma$. Then there exists an $\epsilon > 0$ depending on the geometry of the problem such that the condition numbers of the Hessians are uniformly bounded: $\text{cond}_2(H(y_k)) \leq 1/\epsilon$ for all $k \geq 0$.

To prove this lemma we first define a condition number for a bounded convex set M (the “flatness” of the set M). Using the two-sided ellipsoidal approximation of the set $P(\lambda)$ (by the matrix $H(x(\lambda))$) we then obtain a bound on the condition number of $H(x)$ for x near the path of centers. The proof is given in the Appendix.

■

Note: The bound $1/\epsilon$ on the condition numbers in the preceding Lemma depends on the geometry of the problem (on the “flatness” of the sets $P(\lambda)$), and unfortunately, as simple examples show, the magnitude of the bound of the condition numbers may be as bad as order 2^L , where L is the length of the input of the problem.

There are nondegenerate examples with nonlinear (e.g. convex quadratic) constraint functions for which no such bound exists.

2.7. Extensions

Nesterov and Nemirovsky [12] present an extension of the method presented above handles certain non-convex functions f_i whose level sets $f_i(x) \leq 0$ describe convex domains; for example, the function $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ defined by $f(x, t) := \|x\|_2^2 - t^2$ for $x \in \mathbb{R}^n$, $t \in \mathbb{R}$ and $t > \|x\|_2^2$.

They consider the case that the functions f_i are not necessarily convex, but their barrier functions $-\ln(-f_i(x))$ are self-concordant (and hence convex) with

the additional property that there exists a $\vartheta < \infty$ such that the Newton step h starting at a point $y \in P^o$ for finding the center \bar{x} of P has a length $\|h\|_{H(y)} \leq \vartheta$ uniformly bounded for all $y \in P^o$. Some of the results presented above—like the outer ellipsoidal approximation of P or equation (2.4)—no longer hold, but the convergence of Newton's method for finding the center is the same and convergence of a modified barrier method can be maintained as well.

Another modification of the method is an acceleration when following the smooth path of analytic centers $x(\lambda)$ for $\lambda > \lambda^*$. The tangent to this path can be computed (as $H(y)^{-1}c$) and used as a predictor for a next iterate down the path of centers, while only one or two steps of Newton's method (with line search) will serve as a corrector. Finding the right compromise of staying close enough to the central curve on the one hand and taking large steps along the tangent on the other hand, along with efficient factorizations (or preconditioners) of the matrices $H(y)$ are crucial for a practical program. Implementations of such predictor-corrector type approaches are promising; see e.g. [6, 8, 7, 10].

2.8. Concluding Remarks

There are some difficulties when trying to deduce statements about polynomiality from the above method.

2.8.1. Irrational solutions

The r/NP -model for classifying the “difficulty” of classes of problems is unsatisfactory if one considers interior-point methods that give exactly the same (theoretical) rate of convergence for linear and quadratically constrained convex problems. For linear problems this rate of convergence implies polynomiality of the class of linear programming problems, since one can round the exact (rational) solution from a sufficiently accurate approximation in polynomial time. For the class of convex quadratic problems, no statement about polynomiality can be deduced from this convergence (since a quadratic problem may have an irrational optimal solution that to date cannot be computed by rounding techniques). It is appropriate therefore in a more general context to define the notion of *generalized polynomiality* for a class K of problems if one is able to compute the exact solution of any problem in K up to d digits accuracy in a time that is bounded by a polynomial in d multiplied by a polynomial in the length of the data of the problem.

The definition of generalized polynomiality extends the notion of polynomial-time algorithms in a natural way to problems that do not necessarily have a rational solution. So far such problems have escaped any classification, since the exact solution often could not be computed at all, even if there was a good algorithm to approximate it.

Clearly any problem that is polynomial is also generalized polynomial, and vice versa: a generalized polynomial problem that has a unique rational solution whose length is bounded by a polynomial in the length of the data is also polynomial in the classical sense.

2.8.2. Non-algebraic functions

Concerning the class of α -self-concordant problems, one further difficulty in extending the model of polynomiality is that the "length" of the input cannot be measured in a natural way if the input includes non-algebraic functions.

Acknowledgements

The author would like to thank Prof. M. Saunders for his warm hospitality in Stanford and for his editorial assistance in setting up this paper.

3. APPENDIX

The Appendix is divided into two subsections. In Subsection 3.1 we state some useful and general results. In Subsection 3.2 we present the proofs to which we referred to in Sections 1 and 2.

3.1. Some Useful Lemmas

We begin in recalling a slightly generalized version of the well known Cauchy-Schwarz inequality.

Generalized Cauchy-Schwarz inequality:

If A, M are symmetric matrices with $|x^T M x| \leq x^T A x \quad \forall x \in \mathbb{R}^n$, then also

$$(a^T M b)^2 \leq a^T A a \ b^T A b \quad \forall a, b \in \mathbb{R}^n. \quad (3.1)$$

Proof: Without loss of generality assume that A is positive definite. (Else $A_\epsilon := A + \epsilon I$ is positive definite $\forall \epsilon > 0$, take the limit as $\epsilon \rightarrow 0$ for fixed a, b .) Assume further that $a, b \neq 0$ and set $\mu := \sqrt[4]{\frac{a^T A a}{b^T A b}}$, then it follows from $a^T M b = \frac{1}{4}((a+b)^T M(a+b) - (a-b)^T M(a-b))$ that

$$\begin{aligned} (a^T M b)^2 &\leq \frac{1}{16}((a+b)^T M(a+b) - (a-b)^T M(a-b))^2 \\ &\leq \frac{1}{16}((a+b)^T A(a+b) + (a-b)^T A(a-b))^2 \\ &= \frac{1}{16}(2a^T A a + 2b^T A b)^2 = \frac{1}{4}(a^T A a + b^T A b)^2. \end{aligned}$$

When replacing a by a/μ , and b by μb this implies

$$\begin{aligned} (a^T M b)^2 &= ((\frac{a}{\mu})^T M(\mu b))^2 \leq \frac{1}{4}(\frac{1}{\mu^2} a^T A a + \mu^2 b^T A b)^2 \\ &= (a^T A a)(b^T A b). \end{aligned}$$

■

The following estimate about the spectral radius for symmetric trilinear forms was observed (without proof) by [12].

Spectral radius for symmetric trilinear forms:

If $M \in \mathbb{R}^{n \times n \times n}$ represents a symmetric trilinear form $M : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ and $A \in \mathbb{R}^{n \times n}$ a symmetric bilinear form, and $\mu > 0$ is a scalar such that

$$M[h, h, h]^2 \leq \mu A[h, h]^3 \quad \forall h \in \mathbb{R}^n,$$

then also

$$M[x, y, z]^2 \leq \mu A[x, x]A[y, y]A[z, z] \quad \forall x, y, z \in \mathbb{R}^n. \quad (3.2)$$

Proof: For $x \in \mathbb{R}^n$ denote by M_x the (symmetric) matrix defined by $y^T M_x z := M_x[y, z] := M[x, y, z] \quad \forall y, z \in \mathbb{R}^n$. Without loss of generality let $\mu = 1$ (else substitute A by $\sqrt[3]{\mu} A$). As in the proof of (3.1) assume again that A is positive definite.

By substituting $\tilde{M}[x, y, z] := M[A^{-1/2}x, A^{-1/2}y, A^{-1/2}z]$ one can further assume that $A = I$ is the identity. Finally, it is sufficient to show that

$$|M[x, h, h]| \leq \|x\|_2 \|h\|_2^2 \quad \forall x, h \in \mathbb{R}^n$$

holds, provided that $M[h, h, h]^2 \leq \|h\|_2^6 \quad \forall h \in \mathbb{R}^n$ is true. (The remaining part follows by applying the generalized Cauchy-Schwarz inequality (3.1) for fixed x to M_x !) Let

$$\mu := \max\{M[x, h, h] \mid s.t. \|x\|_2 = \|h\|_2 = 1\}$$

and let \tilde{x}, \tilde{h} be the (not necessarily unique) corresponding arguments. The necessary conditions for a maximum (or a minimum if $M[\tilde{x}, \tilde{h}, \tilde{h}]$ is negative) imply that

$$\begin{pmatrix} M_{\tilde{h}} \tilde{h} \\ 2M_{\tilde{h}} \tilde{x} \end{pmatrix} = \beta \begin{pmatrix} 2\tilde{x} \\ 0 \end{pmatrix} + \rho \begin{pmatrix} 0 \\ 2\tilde{h} \end{pmatrix}$$

where β and ρ are the Lagrange multipliers. From this we deduce that $\beta = \mu/2$ and $\rho = \mu$ (by multiplying from left with $(\tilde{x}^T, \tilde{h}^T)$) and therefore

$$M_{\tilde{h}}(\tilde{x} + \tilde{h}) = \mu(\tilde{x} + \tilde{h}),$$

which also shows that $M[\tilde{h}, \frac{\tilde{x} + \tilde{h}}{\|\tilde{x} + \tilde{h}\|_2}, \frac{\tilde{x} + \tilde{h}}{\|\tilde{x} + \tilde{h}\|_2}] = \mu$. Starting from a maximizing triple $(\tilde{x}, \tilde{h}, \tilde{h})$ this gives a way of generating an(other) maximizing triple $(\tilde{h}, \frac{\tilde{x} + \tilde{h}}{\|\tilde{x} + \tilde{h}\|_2}, \frac{\tilde{x} + \tilde{h}}{\|\tilde{x} + \tilde{h}\|_2})$. Iterating this generating process, one obtains a sequence of maximizing triples that converge³ to a triple $(\gamma(\tilde{x} + \beta\tilde{h}), \gamma(\tilde{x} + \beta\tilde{h}), \gamma(\tilde{x} + \beta\tilde{h}))$. By continuity of M this triple is also maximizing. By assumption however, $M[\gamma(\tilde{x} + \beta\tilde{h}), \gamma(\tilde{x} + \beta\tilde{h}), \gamma(\tilde{x} + \beta\tilde{h})]^2 \leq \|\gamma(\tilde{x} + \beta\tilde{h})\|^6$, which finishes the proof. ■

In the following a quantitative result about the relationship of condition of the Hessian matrix of φ and the shape of the sets $P(\lambda)$ is stated. For this purpose it is useful to define a condition number for the sets $P(\lambda)$.

Definition

Let M be a bounded convex set in \mathbb{R}^n that contains at least two points, and let \bar{M} be its closure. The function $l : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by $l(y) := \max\{y^T(a - b)/\|y\|_2 \mid a, b \in \bar{M}\}$ measures the length of M in direction y . The condition number $\text{cond}_2(M) \in [1, \infty]$ is then defined by

$$r_{\max} := \max\{l(y) \mid \|y\|_2 = 1\}, \quad r_{\min} := \min\{l(y) \mid \|y\|_2 = 1\}, \quad \text{cond}_2(M) := \frac{r_{\max}}{r_{\min}}$$

and is a measure for the “flatness” of M . If M has an interior point, then its condition is finite.

³The establishment of convergence is straightforward. Assume $\tilde{x}^T \tilde{h} = \theta \geq 0$ (else replace \tilde{h} by $-\tilde{h}$). Define $z^{(1)} = \tilde{x}$, $z^{(2)} = \tilde{h}$ and $z^{(k+1)} := (z^{(k)} + z^{(k-1)})/\|(z^{(k)} + z^{(k-1)})\|_2$. We show that $z^{(k)}$ converges to $\gamma(\tilde{x} + \beta\tilde{h})$. Writing $z^{(k)}$ as $\gamma_k(\tilde{x} + \beta_k\tilde{h})$ it follows by induction that $\beta_k \geq 0$ and $\gamma_k \in [\frac{1}{2}, 1]$ (since $\theta \geq 0!$). Computing $\beta_{k+2} = \beta_{k+1} + (\beta_k - \beta_{k+1})\frac{\gamma_1}{\gamma_2 + \gamma_1}$ shows that β_k is a linearly converging sequence. Hence, γ_k converges also. ■

(Note that $r_{\max} = \max\{\|a - b\|_2 \mid a, b \in M\}$ and $r_{\min} = \min\{\|y\|_2 \mid a + y \notin M^\circ \forall a \in M\}.$)

Proposition

Let $M \subset \mathbb{R}^n$ be convex, $\bar{x} \in M$, $H \in \mathbb{R}^{n \times n}$ be positive definite and $\gamma \in \mathbb{R}$ be such that

$$\bar{x} + h \in M \text{ whenever } \|h\|_H < 1 \text{ and } \bar{x} + h \notin M \text{ whenever } \|h\|_H > \gamma.$$

Then

$$\frac{1}{\gamma} \operatorname{cond}_2(M) \leq \sqrt{\operatorname{cond}_2(H)} \leq \gamma \operatorname{cond}_2(M).$$

Proof: Denote by $\nu_1 \leq \nu_2 \dots \leq \nu_n$ the eigenvalues of H . For positive definite H the condition $\operatorname{cond}_2(H)$ is given by $\frac{\nu_n}{\nu_1}$. Using the definition of r_{\max} and r_{\min} and the ellipsoidal approximation of M it is straightforward to show that

$$\frac{2}{\sqrt{\nu_n}} \leq r_{\min} \leq \frac{2\gamma}{\sqrt{\nu_n}} \quad \text{and} \quad \frac{2}{\sqrt{\nu_1}} \leq r_{\max} \leq \frac{2\gamma}{\sqrt{\nu_1}}.$$

From this the claim follows. ■

3.2. Proofs from Chapters 1 and 2

3.2.1. Proof of remark 2

Let the function f fulfill the Relative Lipschitz Condition (1.2) in y and let $f(y) < 0$. Considering the Lagrange remainder formula for the function $g : \mathbb{R} \rightarrow \mathbb{R}$, $g(\theta) := f(y + \theta h)$ we obtain for $\|h\|_{H(y)} \leq 0.5/(1 + M^{1/3})$ that

$$f(y + h) = f(y) + Df(y)h + \frac{1}{2}h^T D^2 f(y + \mu h)h$$

with $\mu \in (0, 1)$.

Suppose now that $f(y + h) \geq 0$, then we have

$$\begin{aligned} 0 < -f(y) &\leq f(y + h) - f(y) = Df(y)h + \frac{1}{2}h^T D^2 f(y + \mu h)h \\ &\leq Df(y)h + \frac{1}{2}h^T D^2 f(y)h(1 + M\mu\|h\|_{H(y)}) \quad (\text{Rel. Lips. Cond.}) \\ &\leq Df(y)h + \frac{1}{2}h^T D^2 f(y)h(1 + \mu \frac{M}{2(1 + M^{1/3})}). \end{aligned}$$

From

$$\left(\frac{Df(y)h}{-f(y)}\right)^2 + \frac{h^T D^2 f(y)h}{-f(y)} = \|h\|_{H(y)}^2 \leq \frac{1}{4(1 + M^{1/3})^2}$$

follows further that

$$Df(y)h \leq \frac{-f(y)}{2(1 + M^{1/3})} \quad \text{and} \quad \frac{h^T D^2 f(y)h}{-f(y)} \leq \frac{-f(y)}{4(1 + M^{1/3})^2}.$$

Substituting this into the first inequality we obtain

$$\begin{aligned} 0 < -f(y) &\leq \frac{-f(y)}{2(1+M^{1/3})} + \frac{1}{2} \frac{-f(y)}{4(1+M^{1/3})^2} \left(1 + \mu \frac{M}{2(1+M^{1/3})}\right) \\ &= -f(y) \left(\frac{1}{2(1+M^{1/3})} + \frac{1}{8(1+M^{1/3})^2} + \frac{\mu}{16} \frac{M}{(1+M^{1/3})^3}\right) < -f(y), \end{aligned}$$

which is a contradiction.

So $f(y+h)$ must be negative as well. ■

3.2.2. Proof of Lemma 1

Suppose a function f is three times continuously differentiable in a point $y \in P^\circ$ and fulfills the Relative Lipschitz Condition (1.2) in y . We verify the self-concordance for an arbitrary fixed direction $h \in \mathbb{R}^n$. Using (1.2) one can bound the function $g : \mathbb{R} \rightarrow \mathbb{R}$,

$$g(\theta) := D^2 f(y)[h, h] - D^2 f(y + \theta h)[h, h]$$

by

$$|g(\theta)| \leq M \|\theta h\|_{H(y)} D^2 f(y)[h, h]$$

for sufficiently small θh . The definition of $\|\cdot\|_{H(y)}$ allows to continue

$$\begin{aligned} |g(\theta)| &\leq M |\theta| \left(\frac{D^2 f(y)[h, h]}{-f(y)} + \left(\frac{D f(y)[h]}{-f(y)} \right)^2 \right)^{1/2} D^2 f(y)[h, h] \\ &\leq M |\theta| \left(\frac{D^2 f(y)[h, h]^{3/2}}{(-f(y))^{1/2}} + \frac{D^2 f(y)[h, h] |D f(y)[h]|}{-f(y)} \right). \end{aligned} \quad (3.3)$$

(Using that $\sqrt{a} + \sqrt{b} \geq \sqrt{a+b}$ for $a, b \geq 0$.) Since $g'(\theta) = -D^3 f(y + \theta h)[h, h, h]$ it follows

$$g(\theta) = g(0) + \theta g'(\mu\theta) = -\theta D^3 f(y + \mu\theta h)[h, h, h]$$

for some $\mu \in (0, 1)$. Hence

$$|g(\theta)| = |\theta D^3 f(y + \mu\theta h)[h, h, h]|.$$

Substituting this into inequality (3.3) yields

$$|D^3 f(y + \mu\theta h)[h, h, h]| \leq M \left(\frac{D^2 f(y)[h, h]^{3/2}}{(-f(y))^{1/2}} + \frac{D^2 f(y)[h, h] |D f(y)[h]|}{-f(y)} \right). \quad (3.4)$$

In the sequel it is helpful to abbreviate the quantities

$$d_1 := \frac{D f(y)[h]}{-f(y)}, \quad d_2 := \frac{D^2 f(y)[h, h]}{-f(y)} \quad \text{and} \quad d_3 := \frac{D^3 f(y)[h, h, h]}{-f(y)}.$$

Without loss of generality assume $d_1 \geq 0$ (otherwise substitute h by $-h$). By convexity of f also $d_2 \geq 0$. Inequality (3.4) is true for sufficiently small θh and some

$\mu = \mu(\theta) \in (0, 1)$. Deviding (3.4) by $-f(y) > 0$ and taking the limit $\theta \rightarrow 0$ we obtain

$$|d_3| \leq M(d_2^{3/2} + d_2 d_1). \quad (3.5)$$

Self-concordance is defined by the derivatives of $\varphi(y) := -\ln(-f(y))$ in (1.1). Observe that

$$D\varphi(y)[h] = d_1, \quad D^2\varphi(y)[h, h] = d_2 + d_1^2 \quad \text{and} \quad D^3\varphi(y)[h, h, h] = d_3 + 3d_2 d_1 + 2d_1^3.$$

Using (3.5) we estimate

$$|D^3\varphi(y)[h, h, h]| \leq |d_3| + 3d_1 d_2 + 2d_1^3 \leq M d_2^{3/2} + (M + 3)d_1 d_2 + 2d_1^3.$$

Comparing this with $(D^2\varphi[y, h])^{3/2} = (d_2 + d_1^2)\sqrt{d_2 + d_1^2}$ it becomes obvious that a suitable multiple of $(D^2\varphi[y, h])^{3/2}$ upper bounds $|D^3\varphi(y)[h, h, h]|$, but finding the best possible multiple is a tedious work which we would like to banish to a footnote⁴. Recalling definition (1.1) we see that $\sqrt{\alpha} = 1 + M$ gives precisely the inequality in the footnote. ■

3.2.3. Proof of Lemma 2

For the sake of completeness we state this proof which is already given in Theorem 1.1 in [12] in slightly modified form.

Let an α -self-concordant function φ , a point $y \in P^\circ$, the gradient $D\varphi(y) = g(y)$, the Hessian matrix $D^2\varphi(y) = H(y)$, an arbitrary vector $h \in \mathbb{R}^n$ with $\delta = \|h\|_{H(y)} < \frac{1}{\sqrt{\alpha}}$ and an arbitrary vector $z \in \mathbb{R}^n$ be given.

Let $s \in [0, 1]$ be such that $y + sh \in P^\circ$. We first show that for such s the inequality

$$(1 - s\delta\sqrt{\alpha})\|z\|_{H(y)} \leq \|z\|_{H(y+sh)} \leq \frac{1}{1 - s\delta\sqrt{\alpha}}\|z\|_{H(y)} \quad (3.6)$$

holds. In a second step one can then show that for $s = 1$ still $y + sh \in P^\circ$. To evaluate how the H -norm of the vectors h and z changes for different matrices $H(y + \rho h)$, with $\rho \in [0, s]$ let us define

$$\Gamma(\rho) := \|h\|_{H(y+\rho h)}^2 = h^T D^2\varphi(y + \rho h)h \geq 0 \quad \text{and}$$

$$\Phi(\rho) := \|z\|_{H(y+\rho h)}^2 = z^T D^2\varphi(y + \rho h)z \geq 0.$$

⁴Abbreviating again $a = d_1$, $b = \sqrt{d_2}$ we obtain

$$\begin{aligned} |D^3\varphi(y)[h, h, h]|^2 &\leq (Mb^3 + Mab^2 + 3ab^2 + 2a^3)^2 \\ &= M^2b^6 + 2M^2ab^5 + 6Mab^5 + 4Ma^3b^3 + M^2a^2b^4 + 6Ma^2b^4 + 4Ma^4b^2 + 9a^2b^4 + 12a^4b^2 + 4a^6 \\ &\text{Using that } 2ab \leq a^2 + b^2 \text{ we eliminate all odd powers and summarize} \\ &\leq 2M^2b^6 + 2M^2a^2b^4 + 11Ma^2b^4 + 3Mb^6 + 6Ma^4b^2 + 9a^2b^4 + 12a^4b^2 + 4a^6. \\ &\leq (4 + 8M + 4M^2)(a^6 + 3a^4b^2 + 3a^2b^4 + b^6) = 4(1 + M)^2(a^2 + b^2)^3 = 4(1 + M)^2(D^2\varphi[h, h])^3 \end{aligned}$$

Summarizing and taking square roots we get $|D^3\varphi(y)[h, h, h]| \leq 2(1 + M)(D^2\varphi(y)[h, h])^{3/2}$.

(Actually, even the constant $2(1 + \frac{1}{\sqrt{2}}M)$ would work.)

In order to show (3.6) we will show that the function Φ is “nearly constant”. The changes of Γ and Φ can be estimated by their derivatives $\Gamma'(\rho)$ and $\Phi'(\rho)$ using the estimate about the spectral radius for symmetric trilinear forms (3.2) proved earlier: From the α -self-concordance of φ follows with (3.2) that

$$|D^3\varphi(y)[z_1, z_2, z_3]| \leq 2\sqrt{\alpha} D^2\varphi(y)[z_1, z_1]^{1/2} D^2\varphi(y)[z_2, z_2]^{1/2} D^2\varphi(y)[z_3, z_3]^{1/2}$$

which implies that

$$|\Gamma'(\rho)| \leq 2\sqrt{\alpha} (h^T D^2\varphi(y + \rho h)h)^{3/2} = 2\sqrt{\alpha} \Gamma(\rho)^{3/2} \quad \text{and}$$

$$|\Phi'(\rho)| \leq 2\sqrt{\alpha} (h^T D^2\varphi(y + \rho h)h)^{1/2} (z^T D^2\varphi(y + \rho h)z) = 2\sqrt{\alpha} \Gamma(\rho)^{1/2} \Phi(\rho).$$

Using the first inequality one can show that $\Gamma^{1/2}$ is “small”, and with the second inequality this implies that $|\Phi'|$ is “small”. There are two cases:

1. $\Gamma(\rho_0) = 0$ for some $\rho_0 \in [0, s]$. This implies $\Gamma(\rho) = 0$ for all $\rho \in [0, s]$ (by integrating $\int_{\rho_0}^{\rho_0+\epsilon} \Gamma'(\rho)d\rho$ for small $|\epsilon|$ and using the first inequality) and then $\Phi'(\rho) = 0$ for all $\rho \in [0, s]$ and $\Phi(s) = \Phi(0)$ which implies that the H -norm of z does not change at all and that (3.6) is true.
2. $\Gamma(\rho) > 0 \quad \forall \rho \in [0, s]$. In this case one can bound $\Gamma^{1/2}(\rho)$ as follows:
 $|\frac{d}{d\rho}(\Gamma(\rho)^{-1/2})| = |\frac{1}{2}\Gamma(\rho)^{-3/2}\Gamma'(\rho)| \leq \sqrt{\alpha} \quad \forall \rho \in [0, s]$ which implies that
 $\Gamma^{-1/2}(\rho) \geq \Gamma^{-1/2}(0) - \rho\sqrt{\alpha} = \frac{1}{\delta} - \rho\sqrt{\alpha} > 0$ (by definition of δ) or that
 $\Gamma^{1/2}(\rho) \leq \delta/(1 - \rho\delta\sqrt{\alpha})$. Inserting this in the second inequality one obtains

$$|\Phi'(\rho)| \leq \frac{2\delta\sqrt{\alpha}}{1 - \rho\delta\sqrt{\alpha}} \Phi(\rho).$$

Again one may conclude (like above) that either $\Phi(\rho) = 0$ on $[0, s]$ (in which case there is nothing to show) or $\Phi(\rho) > 0$ on $[0, s]$. If $\Phi(\rho) > 0$ one can estimate

$$|(\ln \Phi(\rho))'| = \left| \frac{d}{d\rho} \ln(\Phi(\rho)) \right| \leq \frac{2\delta\sqrt{\alpha}}{1 - \rho\delta\sqrt{\alpha}} \quad \text{and thus}$$

$$\begin{aligned} \left| \ln\left(\frac{\Phi(s)}{\Phi(0)}\right) \right| &= \left| \ln \Phi(s) - \ln \Phi(0) \right| = \left| \int_0^s (\ln \Phi(\rho))' d\rho \right| \leq \int_0^s \frac{2\delta\sqrt{\alpha}}{1 - \rho\delta\sqrt{\alpha}} d\rho \\ &= -2\ln(1 - \rho\delta\sqrt{\alpha}) \Big|_0^s = 2\ln\left(\frac{1}{1 - s\delta\sqrt{\alpha}}\right). \end{aligned}$$

This implies that

$$\left(\frac{\Phi(s)}{\Phi(0)}\right)^{1/2} \leq \frac{1}{1 - s\delta\sqrt{\alpha}} \quad \text{and} \quad \left(\frac{\Phi(0)}{\Phi(s)}\right)^{1/2} \leq \frac{1}{1 - s\delta\sqrt{\alpha}}$$

which is inequality (3.6).

A short proof by contradiction shows that $s = 1$ is possible, i.e. that $1 = \max\{\rho \in [0, 1] \mid y + \rho h \in P^\circ\}$: Suppose on the contrary that $1 > s := \sup\{\rho \in [0, 1] \mid y + \rho h \in P^\circ\}$, then by the inequality (3.6) it holds that $D^2\varphi(y + \rho h)$ is bounded for all $\rho \in [0, s]$, and thus $\varphi(y + \rho h)$ is bounded for all $\rho \in [0, s]$. The strong self-concordance of φ implies that $\varphi(x)$ goes to infinity as x approaches the boundary of P , $\lim_{x \rightarrow \partial P} \varphi(x) = \infty$ so that $y + sh \notin \partial P$, the contradiction we were looking for. ■

3.2.4. Proof of Remark 3

Substituting z by h , the Relative Lipschitz condition reduces to

$$|h^T(D^2\varphi(x + h) - D^2\varphi(x))h| \leq \beta(h^T D^2\varphi(x)h)^{3/2}.$$

Defining $\mu(t) := h^T(D^2\varphi(x + th) - D^2\varphi(x))h$, $\mu'(t) = D^3\varphi(x + th)[h, h, h]$, one obtains from $\mu(t) \leq \beta(h^T D^2\varphi(x)h)^{3/2}t$ for $t \geq 0$ that $\mu'(0) \leq \beta(h^T D^2\varphi(x)h)^{3/2}$. This is exactly the condition for α -self-concordance with $\beta = 2\sqrt{\alpha}$ from which the claim follows. ■

3.2.5. Proof of concavity in (2.2)

The proof of concavity of Ψ given in [14] before statement (2.8) can be generalized in a straightforward way to nonlinear convex functions f_i : The term $(\lambda - f_0(x))^q$ has the same structure as the remaining m terms and is therefore omitted here for the sake of clarity. One obtains

$$\frac{D\Psi(x)}{\Psi(x)} = D\ln\Psi(x) = \frac{1}{m} \sum_{i=1}^m \frac{Df_i(x)}{f_i(x)}$$

and

$$\frac{D^2\Psi(x)}{\Psi(x)} - \frac{D^T\Psi(x)D\Psi(x)}{\Psi^2(x)} = D^2\ln\Psi(x) = \frac{1}{m} \sum_{i=1}^m \frac{D^2f_i(x)}{f_i(x)} - \frac{D^Tf_i(x)Df_i(x)}{f_i^2(x)}$$

Hence,

$$\frac{D^2\Psi(x)}{\Psi(x)} = \frac{1}{m} \sum_{i=1}^m \frac{D^2f_i(x)}{f_i(x)} - \frac{D^Tf_i(x)Df_i(x)}{f_i^2(x)} + \frac{1}{m^2} \left(\sum_{i=1}^m \frac{D^Tf_i(x)}{f_i(x)} \right) \left(\sum_{i=1}^m \frac{Df_i(x)}{f_i(x)} \right).$$

Note that for arbitrary vectors h and $v_1, \dots, v_m \in \mathbb{R}^n$ we have

$$h^T(m \sum_i v_i v_i^T)h = m \sum_i (v_i^T h)^2 \geq (\sum_i v_i^T h)^2 = h^T((\sum_i v_i)(\sum_i v_i^T))h.$$

Taking $v_i := D^T f_i(x)/f_i(x)$ and observing that $f_i(x) < 0$ and $\Psi(x) > 0$ this implies that $D^2\Psi(x)$ is negative semidefinite i.e. Ψ is concave. ■

3.2.6. Proof of Lemma 3

This proof proceeds in two steps. First we show that the function φ is well approximated by its quadratic Taylor approximation q . In the second step this information is used to relate the (ellipsoidal) level sets of q to the level sets of $\Psi = e^{-\varphi/m}$. Let $y \in P^\circ$ be arbitrary and define the quadratic approximation q_y of φ in y by

$$q_y(x) := \varphi(y) + D\varphi(y)(x - y) + \frac{1}{2}(x - y)^T D^2\varphi(y)(x - y).$$

For $h \in \mathbb{R}^n$ and sufficiently small μ (such that $\|\mu h\|_{H(y)} \leq 1/\sqrt{\alpha}$) define the difference of φ and q_y in the point $y + \mu h$ by

$$d(\mu) := q_y(y + \mu h) - \varphi(y + \mu h).$$

The Lagrange remainder formula applied to d yields

$$d(\mu) = \frac{\mu^3}{6} d'''(\nu\mu) \quad \text{with some } \nu \in (0, 1).$$

Using the definition of self-concordance one obtains

$$d'''(\mu) = \frac{d^3\varphi(y + \mu h)}{d\mu^3} = D^3\varphi(y + \mu h)[h, h, h] \leq 2\sqrt{\alpha} D^2\varphi(y + \mu h)[h, h]^{3/2}.$$

For $\|\mu h\|_{H(y)} \leq \frac{1}{\sqrt{\alpha}}$ this can further be bounded by Lemma 2:

$$D^2\varphi(y + \mu h)[h, h]^{3/2} \leq \left(D^2\varphi(y)[h, h] \frac{1}{(1 - \sqrt{\alpha}\|\mu h\|_{H(y)})^2} \right)^{3/2}.$$

Inserting the last two estimates in the above Lagrange remainder formula allows us to continue

$$d(\mu) \leq \frac{\mu^3}{6} 2\sqrt{\alpha} \left(D^2\varphi(y)[h, h] \frac{1}{(1 - \sqrt{\alpha}\|\mu h\|_{H(y)})^2} \right)^{3/2} \leq \sqrt{\alpha} \frac{\mu^3}{3} \frac{\|h\|_{H(y)}^3}{(1 - \sqrt{\alpha}\|\mu h\|_{H(y)})^3}$$

for $\|\mu h\|_{H(y)} < 1/\sqrt{\alpha}$. This completes the first step of the proof.

The last inequality will now be used to obtain information about the increase of φ – and thus also about the decrease of Ψ (defined in (2.2)) – around its maximum \bar{x} (the analytic center of P). This allows to construct a decreasing linear function on the ray $\bar{x} + \tilde{h}$, $\mu \geq 0$ that bounds the concave function Ψ in $\mu \in [1, \infty)$ from above. Here, for $h \in \mathbb{R}^n$ we define $\tilde{h} := h/(4\sqrt{\alpha}\|h\|_{H(y)})$. The estimate of d for $h = \tilde{h}$ now implies that

$$d(1) \leq \frac{\sqrt{\alpha}}{3} \frac{\|\tilde{h}\|_{H(y)}^3}{(1 - 1/4)^3} \leq \frac{16\|\tilde{h}\|_{H(y)}^2}{81}.$$

Since $q_{\bar{x}}(\bar{x} + \tilde{h}) = \varphi(\bar{x}) + \frac{1}{2}\|\tilde{h}\|_{H(\bar{x})}^2$ it follows that

$$\varphi(\bar{x} + \tilde{h}) \geq \varphi(\bar{x}) + \frac{1}{2}\|\tilde{h}\|_{H(\bar{x})}^2 - \frac{16\|\tilde{h}\|_{H(y)}^2}{81} > \varphi(\bar{x}) + \frac{3}{10}\|\tilde{h}\|_{H(\bar{x})}^2.$$

Using this and the definition of Ψ yields

$$\begin{aligned}\Psi(\bar{x} + \tilde{h}) &= \exp\left(-\frac{\varphi(\bar{x} + \tilde{h})}{m}\right) < \exp\left(-\frac{\varphi(\bar{x})}{m}\right) \exp\left(\frac{-3}{10m} \|\tilde{h}\|_{H(\bar{x})}^2\right) \\ &= \Psi(\bar{x}) \exp\left(\frac{-3}{10m} \|\tilde{h}\|_{H(\bar{x})}^2\right).\end{aligned}$$

Let $z := \frac{\|\tilde{h}\|_{H(\bar{x})}^2}{m} \leq \frac{1}{16}$. Since $\exp(t) \leq 1 + t + \frac{1}{2}t^2$ for $t \leq 0$ one can conclude that

$$\exp\left(-\frac{3}{10}z\right) \leq 1 - \frac{3}{10}z + \frac{9}{200}z^2 \leq 1 - \frac{z}{4},$$

which implies that

$$\Psi(\bar{x} + \tilde{h}) < \Psi(\bar{x})\left(1 - \frac{z}{4}\right).$$

Setting $\Psi(x) := -\infty$ for $x \notin P$, Ψ is well defined and concave everywhere in \mathbb{R}^n .

Now let h be such that $\|h\|_{H(\bar{x})} \geq 16m\sqrt{\alpha} = \frac{4}{z}\|\tilde{h}\|_{H(\bar{x})}$, then we have $\Psi(\bar{x} + h) < 0$, i.e. $\bar{x} + h \notin P$ (since $\Psi(x) \geq 0$ for $x \in P$). ■

3.2.7. Proof of Lemma 4

(Simplified version of the proof in [12], Prop. 1.2, Th. 1.2, 1.3 and 1.4)

Define $y(s) := y + sh$ for $s \in [0, 1]$ where $h = -D^2\varphi(y)D\varphi(y)^T$ is the Newton step starting in y to minimize φ , then by Lemma 2: $y(s) \in P^\circ$ for all $s \in [0, 1]$ and

$$|z^T(D^2\varphi(y) - D^2\varphi(y(s)))z| \leq \left(\frac{1}{(1 - s\sqrt{\alpha}\|h\|_{H(y)})^2} - 1\right)z^T D^2\varphi(y)z.$$

Using the generalized Cauchy-Schwarz inequality (3.1) and defining $\mu := \sqrt{\alpha}\|h\|_{H(y)}$ we obtain

$$\begin{aligned}|\frac{d}{ds}D\varphi(y(s))z - h^T D^2\varphi(y)z| &= |h^T(D^2\varphi(y(s)) - D^2\varphi(y))z| \leq \\ &\leq \left(\frac{1}{(1 - s\mu)^2} - 1\right)\sqrt{z^T D^2\varphi(y)z}\sqrt{h^T D^2\varphi(y)h} = \left(\frac{1}{(1 - s\mu)^2} - 1\right)\|z\|_{H(y)}\frac{\mu}{\sqrt{\alpha}}.\end{aligned}$$

The left hand side is the absolute value of the derivative $\kappa'(s)$ where κ is defined $\kappa(s) := D\varphi(y(s))z - (1 - s)D\varphi(y)z$. By integration, ($\kappa(0) = 0$)! one can thus bound

$$|\kappa(s)| \leq \frac{\|z\|_{H(y)}}{\sqrt{\alpha}}\mu \int_0^s \frac{1}{(1 - t\mu)^2} - 1 dt = \frac{\|z\|_{H(y)}}{\sqrt{\alpha}}\frac{s^2\mu^2}{1 - s\mu}.$$

For $s = 1$, $y(s) = y + h$ this implies

$$|\kappa(1)| = |D\varphi(y + h)z| \leq \frac{\mu^2}{1 - \mu} \frac{\|z\|_{H(y)}}{\sqrt{\alpha}}.$$

Choosing $z = \tilde{h} = -D^2\varphi(y + h)^{-1}D\varphi(y + h)^T$ as the next Newton step one obtains

$$\|\tilde{h}\|_{H(y+h)}^2 = |D\varphi(y + h)\tilde{h}| \leq \frac{\mu^2}{1 - \mu} \frac{\|\tilde{h}\|_{H(y)}}{\sqrt{\alpha}} \leq \frac{\mu^2}{(1 - \mu)^2} \frac{\|\tilde{h}\|_{H(y+h)}}{\sqrt{\alpha}},$$

the last inequality following from Lemma 2. With $\delta := \mu/\sqrt{\alpha}$ the claim follows when deviding the last line by $\|\tilde{h}\|_{H(y+h)}$. ■

3.2.8. Proof of Remark 4

The previous Lemma implies that if $\delta = \delta_0 \leq \frac{1}{9\sqrt{\alpha}}$ and Newton's method starting in $y_0 := y$ is iterated one obtains a sequence of strictly feasible points $y_k = y_{k-1} + h_{k-1}$ for $k \geq 1$ where the norms $\|.\|_{H(y_k)}$ of the Newton steps h_k converge to zero. Defining $\delta_k := \|h_k\|_{H(y_k)}$ and $\gamma := \frac{81\sqrt{\alpha}}{64}$ it follows (again from Lemma 4) that

$$\delta_k \leq \gamma \delta_{k-1}^2 \leq \frac{1}{\gamma} (\gamma \delta_0)^{2k}.$$

Here, the norm δ_0 of the first Newton step h_0 is $\delta_0 = \|h_0\|_{H(y_0)} \leq \frac{1}{9\sqrt{\alpha}}$. By Lemma 2 this implies for any $z \in \mathbb{R}^n$ that

$$\|z\|_{H(y_1)} \leq \frac{1}{1 - \sqrt{\alpha} \delta_0} \|z\|_{H(y_0)} \leq \frac{9}{8} \|z\|_{H(y_0)}.$$

Since δ_k is a decreasing sequence, this relative change in two subsequent norms is always bounded by $\frac{9}{8}$ so that

$$\|h_k\|_{H(y_0)} \leq \frac{1}{\gamma} \left(\frac{9}{8} \gamma^2 \delta_0^2 \right)^k \leq \frac{1}{\gamma} \left(\frac{3^6}{2^{15}} \right)^k \quad \text{and} \quad \sum_{k=1}^{\infty} \|h_k\|_{H(y_0)} \leq \frac{1}{\gamma} \frac{\frac{9}{8} \gamma^2}{1 - \frac{3^6}{2^{15}}} \delta_0^2 < \frac{3}{2} \sqrt{\alpha} \delta_0^2.$$

Since $\lim_{l \rightarrow \infty} \sum_{k \geq l} \|h_k\|_{H(y_0)} = 0$ also $y_l \rightarrow \bar{x}$ and the claim follows. ■

3.2.9. Proof of Remark 5

(i) First let f be a convex C^2 -function, for $f(x) < 0$ be $\varphi(x) := -\ln(-f(x))$ its logarithmic barrier function, $g(x) := D\varphi(x)^T = \frac{Df(x)^T}{-f(x)}$ the gradient of φ and $H(x, \epsilon) := H(x) + \epsilon I = D^2\varphi(x) + \epsilon I = \frac{Df(x)^T Df(x)}{f^2(x)} + \frac{D^2 f(x)}{-f(x)} + \epsilon I$ a perturbed Hessian matrix of φ . Then $H(x, \epsilon)$ is positive definite for all $\epsilon > 0$ and

$$\|g(x)\|_{H^{-1}(x, \epsilon)}^2 = \frac{Df(x)}{-f(x)} \left(\frac{Df(x)^T Df(x)}{f^2(x)} + \frac{D^2 f(x)}{-f(x)} + \epsilon I \right)^{-1} \frac{Df(x)^T}{-f(x)}.$$

To simplify let $v := \frac{Df(x)^T}{-f(x)}$ and $G := \frac{D^2 f(x)}{-f(x)} + \epsilon I$, then G is positive definite and

$$\|g(x)\|_{H^{-1}(x, \epsilon)}^2 = v^T (G + vv^T)^{-1} v = v^T (G^{-1} - \frac{G^{-1}vv^TG^{-1}}{1 + v^TG^{-1}v}) v < 1.$$

(Note again the equality $\|g\|_{H^{-1}}^2 = g^T H^{-1} g = h^T H h = \|h\|_H^2$ for $h = H^{-1}g$.)

(ii) The second part of the proof now follows immediately (taking the limit as $\epsilon \rightarrow 0$) from another “Cauchy-Schwarz-type” inequality stated in Proposition 3.5 of [12] without further comment (or proof).

If $\|g_i\|_{H_i^{-1}}^2 = \mu_i$ for $1 \leq i \leq m$ (with positive definite matrices H_i)

then $\|\sum_i g_i\|_{(\sum H_i)^{-1}}^2 \leq \sum \mu_i$.

Proof: Observe that

$\mu_i = \min\{\mu \geq 0 \mid (g_i^T h)^2 \leq \mu h^T H_i h \forall h \in \mathbb{R}^n\}$. We want to show that

$\tilde{\mu} := \min\{\mu \geq 0 \mid ((\sum g_i^T h)^2 \leq \mu h^T (\sum H_i)h \forall h \in \mathbb{R}^n\}$ fulfills $\tilde{\mu} \leq \sum \mu_i$. We may also write

$\tilde{\mu} = \min\{\mu \geq 0 \mid (\sum g_i^T h)^2 \leq \mu \sum h^T H_i h \forall h : (g_i^T h)^2 \leq \mu_i h^T H_i h\}$ (since the last set of inequalities is by definition of μ_i always satisfied). From this definition of $\tilde{\mu}$ it is obvious that $\bar{\mu} \geq \tilde{\mu}$ for any $\bar{\mu}$ satisfying

$$0 = \sup\{(\sum \alpha_i)^2 - \bar{\mu} \sum \beta_i \mid \alpha_i \in \mathbb{R}, \beta_i \geq 0, \alpha_i^2 - \mu_i \beta_i \leq 0\}.$$

(If we added the additional restriction that $\alpha_i = g_i^T h$ and $\beta_i = h^T H_i h$, then obviously $\bar{\mu} = \tilde{\mu}$ would be feasible, i.e. would be satisfying that the "sup" = 0. By allowing a (possibly) larger set α_i, β_i here, the "sup" may increase and it might require a larger value $\bar{\mu} > \tilde{\mu}$ to ensure feasibility of $\tilde{\mu}$.) For any $\rho \neq 0$ the sign of the "sup" is invariant under the transformation " $\forall i : \beta_i \rightarrow \rho^2 \beta_i, \alpha_i \rightarrow \rho \alpha_i$ ". Hence we may add the additional constraint $\sum \beta_i \leq 1$ while keeping the same set of feasible values $\tilde{\mu}$ and guaranteeing that on the resulting compact domain the "sup" is actually a maximum for which we can consider the necessary conditions. For this purpose define $e := (1, 1, \dots, 1)^T \in \mathbb{R}^m$, $\alpha := (\alpha_1, \alpha_2, \dots, \alpha_m)^T \in \mathbb{R}^m$, $\xi_0 := (0, \dots, 0; e^T)^T \in \mathbb{R}^{2m}$ and

$$\xi_i := (0, \dots, 0, 2\alpha_i, 0, \dots, 0; 0, \dots, 0, -\mu_i, 0, \dots, 0)^T \in \mathbb{R}^{2m}$$

for $1 \leq i \leq m$, where only the i -th and the $(m+i)$ -th entry of ξ_i are nonzero. Since $\beta_i = 0$ implies $\alpha_i = 0$ we may restrict ourselves to $\beta_i > 0$. (And also $\mu_i > 0$.) Then the necessary conditions for a maximum imply

$$(2e^T \alpha \cdot e^T; -\bar{\mu} e^T)^T = \rho_0 \xi_0 + \rho_1 \xi_1 + \dots + \rho_m \xi_m,$$

where $\rho_i \in \mathbb{R}$ are the Lagrange multipliers. For $i \geq 1$ we deduce from the $(m+i)$ -th entry of ξ_i that $\rho_i = \frac{\bar{\mu} + \rho_0}{\mu_i}$, and the i -th entry tells us then that $\alpha_i = e^T \alpha \frac{\mu_i}{\bar{\mu} + \rho_0}$. Substituting this in the "sup" yields

$$0 = \max\left\{\left(\frac{e^T \alpha}{\bar{\mu} + \rho_0}\right)^2 (\sum \mu_i)^2 - \bar{\mu} (\sum \beta_i) \mid \left(\frac{e^T \alpha}{\bar{\mu} + \rho_0} \mu_i\right)^2 \leq \mu_i \beta_i\right\}.$$

Substituting now β_i we may continue

$$\leq \max\left\{\left(\frac{e^T \alpha}{\bar{\mu} + \rho_0}\right)^2 (\sum \mu_i)^2 - \bar{\mu} \sum \left(\frac{e^T \alpha}{\bar{\mu} + \rho_0}\right)^2 \mu_i\right\}.$$

Factoring $\left(\frac{e^T \alpha}{\bar{\mu} + \rho_0}\right)^2 \geq 0$ it is obvious that the last term is zero for any $\bar{\mu} \geq \sum \mu_i$ (and in particular for $\bar{\mu} = \sum \mu_i$). ■

3.2.10. Proof of Remark 6

Let again $H(x) = D^2 \varphi(x)$, $g(x) = D\varphi(x)^T$ and $h(x) = -H(x)^{-1}g(x)$ be the Newton step starting in x for finding the analytic center \bar{x} of P . Let $y \in P^\circ$ and $\epsilon \leq \frac{1}{4}$ be given such that $\|\bar{x} - y\|_{H(\bar{x})} \leq \frac{\epsilon}{\sqrt{\alpha}}$ and set $\bar{h} := \bar{x} - y$.

For some fixed vector $z \in \mathbb{R}^n$ define $l : \mathbb{R} \rightarrow \mathbb{R}$ by $l(\mu) := g(y + \mu\bar{h})^T z$, then $l(1) = 0$ (for any z) and

$$l'(\mu) = \bar{h}^T H(y + \mu\bar{h})z, \quad l''(\mu) = D^3\varphi(y + \mu\bar{h})[\bar{h}, \bar{h}, z],$$

and $l(\mu) = l(0) + \mu l'(0) + \frac{1}{2}\mu^2 l''(\xi)$ for some $\xi \in [0, \mu]$ by the Lagrange remainder formula. For $\mu = 1$ one obtains

$$0 = g(y)^T z + \bar{h}^T H(y)z + \frac{1}{2}D^3\varphi(y + \xi\bar{h})[\bar{h}, \bar{h}, z]$$

From α -self-concordance follows with (3.2) that

$$|z^T(g(y) + H(y)\bar{h})| \leq \sqrt{\alpha} \|\bar{h}\|_{H(y)} \|z\| \leq \sqrt{\alpha} \|\bar{h}\|_{H(y)} \|z^T H(y + \xi\bar{h})z\|^{1/2}.$$

Let $d := g(y) + H(y)\bar{h}$ (then $H(y)^{-1}d = h(y) - \bar{h}$) and let $z := \frac{H(y)^{-1}d}{\|H(y)^{-1}d\|_{H(y)}} = \frac{H(y)^{-1}d}{(d^T H(y)^{-1}d)^{1/2}}$, then the above formula reduces to

$$\frac{d^T H(y)^{-1}d}{(d^T H(y)^{-1}d)^{1/2}} \leq \sqrt{\alpha} \|\bar{h}\|_{H(y+\xi\bar{h})} \frac{\|H(y)^{-1}d\|_{H(y+\xi\bar{h})}}{\|H(y)^{-1}d\|_{H(y)}}. \quad (3.7)$$

By assumption, $\|\bar{h}\|_{H(y+\bar{h})} \leq \frac{\epsilon}{\sqrt{\alpha}}$. Relating the norms $\|\cdot\|_{H(y+\bar{h})}$, $\|\cdot\|_{H(y+\xi\bar{h})}$ and $\|\cdot\|_{H(y)}$ again by lemma 2 it is straightforward to show that for $\epsilon \leq \frac{1}{4}$ inequality (3.7) implies

$$\|H(y)^{-1}d\|_{H(y)} = (d^T H(y)^{-1}d)^{1/2} \leq \frac{1}{\sqrt{\alpha}} \frac{\epsilon^2}{(1-\epsilon)^2}$$

for any $\xi \in [0, 1]$.)⁵ Note that $\bar{x} = H(y)^{-1}d$ is the result of the Newton step. Applying Lemma 2 one more time yields

$$\|H(y)^{-1}d\|_{H(\bar{x})} \leq \frac{1}{\sqrt{\alpha}} \frac{\epsilon^2}{(1-\epsilon)^3}$$

which establishes quadratic convergence. ■

3.2.11. Proof of (2.3)

Suppose y_k satisfies $\delta_k := \|h_k\|_{H(y_k, \lambda_k)} = \|h(y_k, \lambda_k)\|_{H(y_k, \lambda_k)} \leq 1/(4\sqrt{\alpha})$ and $y_{k+1} = y_k + h_k$.

By Lemma 4 then $\tilde{\delta} := \|h(y_{k+1}, \lambda_k)\|_{H(y_{k+1}, \lambda_k)} \leq 1/(9\sqrt{\alpha})$.

We examine the effect on h and H caused by the update of λ_{k+1} . Denote by $g(y, \lambda)$ the gradient

$$g(y, \lambda)^T := L\varphi(y, \lambda) = q \frac{c^T}{(\lambda - c^T y)} + \sum_{i=1}^m \frac{Df_i(y)}{f_i(y)}$$

⁵For $\xi = 0$ or $\xi = 1$ it follows directly from Lemma 2, for $\xi \in (0, 1)$ it follows when applying Lemma 2 twice and using $\epsilon \leq \frac{1}{4}$.

(φ as defined in (2.1)) and $h(y_{k+1}, \lambda_k) = -H(y_{k+1}, \lambda_k)^{-1}g(y_{k+1}, \lambda_k)$ ($\leq 1/(9\sqrt{\alpha})$). The update of $\lambda_{k+1} = \lambda_k - \sigma(\lambda_k - c^T y_{k+1})$ effects

$$g(y_{k+1}, \lambda_{k+1}) = g(y_{k+1}, \lambda_k) + q \frac{\sigma c}{\lambda_{k+1} - c^T y_{k+1}} = g(y_{k+1}, \lambda_k) + q\sigma\tilde{c}$$

with $\tilde{c} = c/(\lambda_{k+1} - c^T y_{k+1})$. The H^{-1} -norm of the second part can be bounded by

$$\begin{aligned} \|q\sigma\tilde{c}\|_{H(y_{k+1}, \lambda_{k+1})^{-1}} &= \left(q\sigma\tilde{c}^T H(y_{k+1}, \lambda_{k+1})^{-1} q\sigma\tilde{c} \right)^{1/2} \\ &\leq \sigma q \left(\tilde{c}^T \left(H(y_{k+1}, \infty) + q\tilde{c}\tilde{c}^T \right)^{-1} \tilde{c} \right)^{1/2} \leq \sigma\sqrt{q} \end{aligned}$$

(since $H(y_{k+1}, \infty)$ is positive definite).

Further, from $H(y_{k+1}, \lambda_{k+1}) = H(y_{k+1}, \lambda_k) + q\sigma^2\tilde{c}\tilde{c}^T$ follows that

$$\begin{aligned} \|h(y_{k+1}, \lambda_{k+1})\|_{H(y_{k+1}, \lambda_{k+1})} &= \|g(y_{k+1}, \lambda_{k+1})\|_{H(y_{k+1}, \lambda_{k+1})^{-1}} \\ &\leq \|g(y_{k+1}, \lambda_k)\|_{H(y_{k+1}, \lambda_{k+1})^{-1}} + \|q\sigma\tilde{c}\|_{H(y_{k+1}, \lambda_{k+1})^{-1}} \\ &\leq \|g(y_{k+1}, \lambda_k)\|_{H(y_{k+1}, \lambda_k)^{-1}} + \sigma\sqrt{q} \leq \frac{1}{9\sqrt{\alpha}} + \sigma\sqrt{q}. \end{aligned}$$

(Here, q is chosen $q = m$.) This shows (2.3) ■

3.2.12. Proof of (2.4)

Denote the analytic center $x(\lambda_k)$ by \bar{x} , then the iterate y_{k+1} meets the assumptions of Remark 4 so that

$$\begin{aligned} \|y_{k+1} - \bar{x}\|_{H(y_{k+1}, \lambda_k)} &\leq \|h(y_{k+1}, \lambda_k)\|_{H(y_{k+1}, \lambda_k)} + \|y_{k+1} + h(y_{k+1}, \lambda_k) - \bar{x}\|_{H(y_{k+1}, \lambda_k)} \\ &\leq \frac{1}{9\sqrt{\alpha}} + \frac{3\sqrt{\alpha}}{2} \left(\frac{1}{9\sqrt{\alpha}} \right)^2 \leq \frac{7}{54\sqrt{\alpha}}. \end{aligned}$$

By the equivalence of the H -norms (Lemma 2) the same distance measured in the central norm fulfills $\|y_{k+1} - \bar{x}\|_{H(\bar{x}, \lambda_k)} \leq \|y_{k+1} - \bar{x}\|_{H(y_{k+1}, \lambda_k)} / (1 - \frac{7}{54}) \leq \frac{0.15}{\sqrt{\alpha}}$. From the inner ellipsoidal approximation of $P(\lambda)$ in part 1 of Lemma 2 follows that y_{k+1} is at most 15 percent “away” from the center,

$$c^T y_{k+1} - c^T \bar{x} \leq 0.15 \left(\max_{x \in P(\lambda_k)} \{c^T x\} - c^T \bar{x} \right) \leq 0.15(\lambda_k - c^T \bar{x}).$$

It is easy to show (see, e.g. [3], Lemma (3.8)) that in the case of convex constraint functions f_i and a linear objective function f_0 the following inequality holds in the analytic center $x(\lambda)$ of $P(\lambda)$:

$$\lambda - c^T x(\lambda) \geq \frac{1}{2}(\lambda - \lambda^*).$$

(Only for $q \geq m$ in (2.1)). With the previous result this implies that

$$\lambda - c^T y_{k+1} \geq \frac{0.85}{2}(\lambda - \lambda^*)$$

from which the claim follows. ■

3.2.13. Proof of Lemma 5

In the beginning of this chapter the condition number of a bounded convex set M has been defined. Under the assumptions of section 1, the set $P = P(\lambda = \infty)$ has a finite condition $\text{cond}_2(P) < \infty$. Since the program is also nondegenerate there exists a $\lambda_0 > \lambda^*$ (the unknown optimal value) such that $P(\lambda_0)$ is a simplex S bounded by the objective function $f_0(x) \leq \lambda_0$ and the n linearly independent constraints that are active in the optimum. For $\lambda \in (\lambda^*, \lambda_0]$ the set $P(\lambda)$ is similar to S and $\text{cond}_2(P(\lambda)) = \text{cond}_2(S)$. It is a simple exercise to verify that for $\lambda > \lambda^*$ the condition $\text{cond}_2(P(\lambda))$ is a continuous function of λ . Since the limits $\lim_{\lambda \rightarrow \infty} \text{cond}_2(P(\lambda)) = \text{cond}_2(P)$ and $\lim_{\lambda \rightarrow \lambda^*} \text{cond}_2(P(\lambda)) = \text{cond}_2(S)$ are finite one may conclude that there exists a number $C < \infty$ such that

$$\text{cond}_2(P(\lambda)) \leq C \text{ for } \lambda \in (\lambda^*, \infty).$$

Remark: In special cases it may happen that for some $\lambda \in (\lambda^*, \infty)$ the condition numbers are not monotone and $\text{cond}_2(P(\lambda)) > \max\{\text{cond}_2(P), \text{cond}_2(S)\}$ it seems however that always $\text{cond}_2(P(\lambda)) \leq \text{cond}_2(P) + \text{cond}_2(S)$ holds.

We recall the ellipsoidal approximation of the sets $P(\lambda)$ around the analytic centers $x(\lambda)$ with the matrices $H(x(\lambda), \lambda)$ and $\gamma = m - 1$. (For the case of a Linear Program [14] proved a better inclusion with similarity ratio $(m - 1)$.)

Therefore $\text{cond}_2(H(x(\lambda), \lambda)) \leq (m - 1)^2 C^2 =: \tilde{C}$ for $\lambda \in (\lambda^*, \infty)$.

Now let a point y_k lie in the domain of quadratic convergence of Newton's method, i.e. $\delta := \|h(y_k, \lambda_k)\|_{H(y_k, \lambda_k)} \leq \frac{1}{4\sqrt{\alpha}}$, where $x(\lambda_k)$ is the "nearest" center. From the proof of (2.4) (and from Lemma 2) follows that the Hessian matrices in $x(\lambda_k)$ and in $y_k + h(y_k, \lambda_k)$ fulfill

$$0.85\|z\|_{H(x(\lambda_k), \lambda_k)} \leq \|z\|_{H(y_k + h(y_k, \lambda_k), \lambda_k)} \leq \frac{1}{0.85}\|z\|_{H(x(\lambda_k), \lambda_k)}$$

for any $z \in \mathbb{R}^n$. Similarly, the Hessian matrices in y_k and in $y_k + h(y_k, \lambda_k)$ fulfill the same relationship with the factors $\frac{3}{4}$ resp. $\frac{4}{3}$ (by Lemma 2). Putting this together the eigenvalues of the Hessian matrices in y_k and $x(\lambda_k)$ change at most by a factor of $\theta \in (\frac{51}{80}, \frac{80}{51})$ so that $\text{cond}_2(H(y_k, \lambda_k)) \leq \frac{80^2}{51^2} \text{cond}_2(H(x(\lambda_k), \lambda_k)) \leq 3\text{cond}_2(H(x(\lambda_k), \lambda_k))$. This completes the proof of Lemma 5.

References

- [1] M. Groetschel, L. Lovasz, S. Schrijver, *Geometric algorithms and combinatorial optimization*, (Springer Verlag, Heidelberg, 1988).
- [2] D. den Hertog, C. Roos, T. Terlaky, "On the classical logarithmic barrier function method for a class of smooth convex programming problems", Report 90-28, Delft University of Technology, The Netherlands (1990).
- [3] F. Jarre, "On the convergence of the method of analytic centers when applied to convex quadratic programs", Report No. 35 (1987), Schwerpunktprogramm der DFG Anwendungsbezogene Optimierung und Steuerung. To appear in *Mathematical Programming* 49 (1991).
- [4] F. Jarre, "The method of analytic centers for solving smooth convex programs", *Lecture Notes in Mathematics*, Vol. 1405 Optimization, S. Dolecki ed., Springer (1989) 69-85.
- [5] F. Jarre, *The method of analytic centers for smooth convex programs*, (Grottenthaler Verlag, Bamberg, 1989).
- [6] F. Jarre, G. Sonnevend, J. Stoer, "An implementation of the method of analytic centers", in: A. Benoussan, J.L. Lions eds., *Lecture Notes in Control and Information Sciences* 111 (New York, Springer 1988).
- [7] I.J. Lustig, R.E. Marsten, D.F. Shanno, "On implementing Mehrotra's predictor corrector interior point method for linear programming", Technical Report SOR 90-03, Dept. of Civil Eng. and OR, Princeton University, Princeton, NJ 08544 (1990).
- [8] S. Mehrotra, "On the implementation of a (primal-dual) interior point method", Technical Report 90-03, Dept. of Ind. Engineering and Management Sciences, Northwestern University, Evanston, IL (1990).
- [9] S. Mehrotra, J. Sun, "An interior point algorithm for solving smooth convex programs based on Newton's method", Technical Report 88-08, Dept. of Ind. Engineering and Management Sciences, Northwestern University, Evanston, IL (1988).
- [10] J. Mennicken, "Implementation of a first order central path following algorithm for solving large linear programs", Report No. 202, Schwerpunktprogramm der DFG Anwendungsbezogene Optimierung und Steuerung, Institut für Ang. Math. und Statistik, Universität Würzburg, Am Hubland (1990).
- [11] J.E. Nesterov A.S. Nemirovsky, "A general approach to polynomial-time algorithms design for convex programming", report, Central Economical and Mathematical Institute, USSR Acad. Sci. (Moscow, USSR, 1988).
- [12] J.E. Nesterov A.S. Nemirovsky, "Self-concordant functions and polynomial-time methods in convex programming", report, Central Economical and Mathematical Institute, USSR Acad. Sci. (Moscow, USSR, 1989).
- [13] J. Renegar, "A polynomial-time algorithm based on Newton's method for linear programming", *Mathematical Programming* 40 (1988) 59-93.
- [14] G. Sonnevend, "An "analytical centre" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming", *Lecture Notes of Control and Information Sciences* 84 (1986) 866-878.
- [15] G. Sonnevend, J. Stoer, "Global ellipsoidal approximations and homotopy methods for solving convex analytic programs", *Appl. Math. and Optimization* 21 (1989) 139-166.
- [16] G. Sonnevend, J. Stoer, G. Zhao, "On the complexity of following the central path by linear extrapolation in linear programs", to appear in U. Rieder and P. Kleinschmidt eds., *Proc. 14 Symp. on Operations Research* (Ulm 1989).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED	
	November 1990	Technical Report	
4. TITLE AND SUBTITLE		5. FUNDING NUMBERS	
Interior-Point Methods for Convex Programming		N00014-90-J-1242	
6. AUTHOR(S)			
Florian Jarre			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)		8. PERFORMING ORGANIZATION REPORT NUMBER	
Department of Operations Research - SOL Stanford University Stanford, CA 94305-4022		1111MA	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
Office of Naval Research - Department of the Navy 800 N. Quincy Street Arlington, VA 22217		SOL 90-16	
12a. DISTRIBUTION AVAILABILITY STATEMENT		12b. DISTRIBUTION CODE	
UNLIMITED		UL	
13. ABSTRACT (Maximum 200 words)			
<p>This work is concerned with generalized convex programming problems, where the objective and also the constraints belong to a certain class of convex functions. It examines the relationship of two conditions for generalized convex programming—self-concordance and a relative Lipschitz condition—and gives an outline for a short and simple analysis of an interior point method for generalized convex programming. It generalizes ellipsoidal approximations for the feasible set, and in the special case of a nondegenerate linear program it establishes a uniform bound on the condition number of the matrices occurring when the iterates remain near the path of centers.</p>			
14. SUBJECT TERMS		15. NUMBER OF PAGES	
convex program; ellipsoidal approximation; relative Lipschitz condition; self concordance.		27 pp.	
		16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED			SAR